PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 14/00

A2

(11) International Publication Number: **WO 00/24774**

(43) International Publication Date: 4 May 2000 (04.05.00)

(21) International Application Number: PCT/US99/25458

(22) International Filing Date: 28 October 1999 (28.10.99)

(30) Priority Data:

09/181,711	28 October 1998 (28.10.98)	US
Not furnished	28 October 1998 (28.10.98)	US
09/209,547	11 December 1998 (11.12.98)	US
Not furnished	11 December 1998 (11.12.98)	US
09/313,457	17 May 1999 (17.05.99)	US
Not furnished	17 May 1999 (17.05.99)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	09/209,547 (CIP)
Filed on	11 December 1998 (11.12.98)
US	09/313,457 (CIP)
Filed on	17 May 1999 (17.05.99)
US	09/181,711 (CIP)
Filed on	28 October 1998 (28.10.98)
US	Not furnished (CIP)
Filed on	28 October 1998 (28.10.98)
US	Not furnished (CIP)
Filed on	11 December 1998 (11.12.98)
US	Not furnished (CIP)
Filed on	17 May 1999 (17.05.99)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). AU–YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GROWTH FACTOR RELATED MOLECULES

```
1   MRGTRLALLALVLA---ACGELAPALRCYV  2777282
1   MASRWAVQLLLVAAWSMGCGE---ALKCYT  GI 1536902

28  CPEPTGVSDCVTIATCTTNETMCKTTLYSR  2777282
28  CKEPMTSASCRTITRCKPEDTACMTTLVTV  GI 1536902

58  EIVYPFQGDSTVTKSCASKCKPSDVDGIGQ  2777282
58  EAEYPFNQSPVVTRSCSSSCVATDPDSIGA  GI 1536902

88  TLPVSCCNTELCNVDGAPALNSLHCGALTL  2777282
88  AHLIFCCFRDLCNSE----------------  GI 1536902

118 LPLLSLRL                        2777282
103 --------L                       GI 1536902
```

(57) Abstract

The invention provides human growth factor related molecules (GFRP) and polynucleotides which identify and encode GFRP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of GFRP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GROWTH FACTOR RELATED MOLECULES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human growth factor related molecules and to the use of these sequences in the diagnosis, treatment, and prevention of developmental disorders; cell proliferative disorders including cancer; immune disorders including inflammation; reproductive and cardiovascular disorders; and infections.

BACKGROUND OF THE INVENTION

Intercellular communication is essential for the development and survival of multicellular organisms. Communication is achieved through the secretion of proteins by signaling cells and the internalization of these proteins by target cells. Growth factors are an example of secreted proteins that mediate communication between signaling and target cells. Inside a signaling cell, growth factors are synthesized and transported through the secretory pathway. Entry into a secretory pathway is mediated by a signal peptide sequence, a protein sorting motif at the N-terminus of most secreted proteins. Within the secretory pathway, the signal sequence is proteolytically removed from its cognate growth factor. Most growth factors also undergo further post-translational modifications within the secretory pathway. These modifications can include glycosylation, phosphorylation, and intramolecular disulfide bond formation. Following secretion into the extracellular space, some growth factors oligomerize or associate with extracellular matrix components. Secreted growth factors bind to specific receptors on the surfaces of their target cells, and the bound receptors trigger second messenger signal transduction pathways. These signal transduction pathways elicit specific cellular responses in target cells. These responses can include the modulation of gene expression and the stimulation or inhibition of cell division, cell differentiation, and cell motility.

Most growth factors are local mediators that act on cells in the immediate environment. Such local activity is maintained by physical proximity of a signaling cell to its target cell, sequestration of the growth factor by extracellular matrix components, internalization and degradation of the growth factor by the target cell, and exclusion of the growth factor from circulation.

Growth factors fall into three broad and overlapping classes. The first and broadest class includes the large polypeptide growth factors, which are wide-ranging in their effects. These factors include epidermal growth factor (EGF), fibroblast growth factor (FGF), transforming growth factor-β (TGF-β), insulin-like growth factor (IGF), nerve growth factor (NGF), and platelet-derived growth factor (PDGF), each defining a family of numerous related factors. The large polypeptide growth factors generally act as mitogens on diverse cell types to stimulate wound healing, bone synthesis and remodeling, extracellular matrix synthesis, and proliferation of epithelial, epidermal, and connective

tissues. Some members of the TGF-β, EGF, and FGF families also function as inductive signals in the differentiation of embryonic tissue. However, some of the large polypeptide growth factors carry out specific functions on a restricted set of target tissues. For example, mouse growth/differentiation factor 9 (GDF-9) is a TGF-β family member that is expressed solely in the ovary (McPherron, A.C. and S.-J. Lee (1993) J. Biol. Chem. 268:3444-3449). NGF functions specifically as a neurotrophic factor, promoting neuronal growth and differentiation.

Follistatin (FS) is a protein that specifically binds and inhibits activin, a member of the transforming growth factor-β family of growth and differentiation factors. Activin performs a variety of functions associated with growth and differentiation, including induction of mesoderm in the developing embryo and regulation of female sex hormone secretion in the adult (de Krester, D.M. (1998) J. Reprod. Immunol. 39:1-12). Both activin and FS are found in many types of cells. The interaction of FS and activin influences a variety of cellular processes in the gonadal tissues, the pituitary gland, membranes associated with pregnancy, the vascular tissues, and the liver (reviewed in Phillips, D.J. and D.M. de Krester (1998) Front. Neuroendocrinol. 19:287-322). FS may also play a direct role in the neuralization of embryonic tissue (Hemmati-Brivanlou et al. (1994) Cell 77:283-295).

FS is conserved among diverse species such as frog, chicken, and human. Variants of human FS include a 288 amino acid and a 315 amino acid isoform (McConnell, D.S. et al. (1998) J. Clin. Endocrinol. Metab. 83:851-858). Most follistatins contain a conserved domain with ten regularly spaced cysteine residues. These residues are likely involved in disulfide bond formation and the binding of cations. Similar domains are observed in Kazal protease inhibitors and osteonectin (also called SPARC or BM-40), an extracellular matrix-associated glycoprotein expressed in a variety of tissues during embryogenesis and repair (reviewed in Lane, T.F. and E.H. Sage (1994) FASEB J. 8:163-173). Osteonectin contains not only an FS-like polycysteine domain, but also other modular domains that can function independently to bind cells and matrix components and can change cell shape by selectively disrupting cellular contacts with matrix. High levels of osteonectin are associated with developing bones and teeth, principally osteoblasts, odontoblasts, and perichondrial fibroblasts of embryos. Osteonectin modulation of cell adhesion and proliferation may also function in tissue remodeling and angiogenesis (Kupprion et al. (1998) J. Biol. Chem. 45:29635-29640).

FS is associated with a variety of cell proliferative, reproductive, and developmental disorders. Transgenic mice lacking FS have multiple musculoskeletal defects and die shortly after birth (Matzuk, M.M. et al. (1995) Nature 374:360-363). Abnormal expression and localization of FS have been implicated in benign prostatic hyperplasia and prostate cancer (Thomas, T.Z. et al. (1998) Prostate 34:34-43). The Follistatin-Related Gene, which encodes a protein with a FS-like polycysteine domain, is associated with chromosomal translocations that may play a role in

leukemogenesis (Hayette, S. (1998) Oncogene 16:2949-2954). In the inflammatory response, FS increases the macrophage foam cell formation characteristic of early atherosclerosis (Kozaki, K. et al. (1997) Arterioscler. Thromb. Vasc. Biol. 17:2389-2394).

The bone morphogenetic proteins (BMPs) are bone-derived factors capable of inducing ectopic bone formation (Wozney, J.M. et al. (1988) Science 242:1528-1534). BMPs are hydrophobic glycoproteins involved in bone generation and regeneration, several of which are related to the TGF-beta superfamily. BMP-1, for example, appears to have a regulatory role in bone formation and is characterized by procollagen C-proteinase activity and the presence of an extracellular "CUB" domain. The CUB domain is composed of some 110 residues containing four cysteines which probably form two disulfide bridges, and is found in a variety of functionally diverse, mostly developmentally regulated proteins (ExPASy PROSITE document PR00908).

The second class of growth factors includes the hematopoietic growth factors, which have a narrow target specificity. These factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and stem cell precursors. These factors include the colony-stimulating factors (e.g., G-CSF, M-CSF, GM-CSF, and CSF1-3), erythropoietin, and the cytokines.

Cytokines comprise a family of signaling molecules that modulate the immune system and the inflammatory response. Cytokines are specialized hematopoietic factors secreted by cells of the immune system, usually leukocytes (white blood cells), in response to external insults, such as tissue damage and viral or microbial infection. However, other tissues are also capable of secreting cytokines in response to disease or trauma. Cytokines function in tissue repair, inflammation, and modulation of the immune response. Cytokines act as growth and differentiation factors that primarily affect cells of the immune system such as B- and T-lymphocytes, monocytes, macrophages, and granulocytes. Like other signaling molecules, cytokines bind to specific plasma membrane receptors and trigger intracellular signal transduction pathways which alter gene expression patterns. There is considerable potential for the use of cytokines in the treatment of inflammation and immune system disorders.

Cytokine structure and function have been extensively characterized in vitro. Most cytokines are small polypeptides of about 30 kilodaltons or less. Over 50 cytokines have been identified from human and rodent sources. Examples of cytokine subfamilies include the interferons (IFN-α, -β, and -γ), the interleukins (IL1-IL13), the tumor necrosis factors (TNF-α and -β), and the chemokines. Many cytokines have been produced using recombinant DNA techniques, and the activities of individual cytokines have been determined in vitro. These activities include regulation of leukocyte proliferation, differentiation, and motility.

The activity of an individual cytokine in vitro may not reflect the full scope of that cytokine's

activity in vivo. Cytokines are not expressed individually in vivo but are instead expressed in combination with a multitude of other cytokines when the organism is challenged with a stimulus. Together, these cytokines collectively modulate the immune response in a manner appropriate for that particular stimulus. Therefore, the physiological activity of a cytokine is determined by the stimulus itself and by complex interactive networks among co-expressed cytokines which may demonstrate both synergistic and antagonistic relationships.

Recently, a unique cytokine has been isolated that appears to have anti-tumor activity in vitro (Ridge, R.J. and N.H. Sloane (1996) Cytokine 8:1-5). This cytokine, anti-neoplastic urinary protein (ANUP), was originally purified as a dimer from human urine. ANUP was later classified as a cytokine when localization studies demonstrated that it was expressed in human granulocytes. ANUP inhibits the growth of cell lines derived from tumors of the breast, skin, lung, bladder, pancreas, and cervix. However, ANUP does not affect the growth of human non-tumor cell lines. The N-terminal 22 amino acids of ANUP comprise a signal peptide which is cleaved from the mature protein. The first nine amino acids of the mature protein retain about 10% of the anti-tumor activity. In addition, ANUP contains a Ly-6/u-PAR sequence motif that is typical of certain cell surface glycoproteins. This motif is characterized by a distinct pattern of six cysteine residues within a 50-residue consensus sequence. The Ly-6/u-PAR motif is found in the Ly-6 T-lymphocyte surface antigen and in the receptor (u-PAR) for urokinase-type plasminogen activator, an extracellular serine protease.

Chemokines comprise a cytokine subfamily with over 30 members. (Reviewed in Wells, T.N.C. and M.C. Peitsch (1997) J. Leukoc. Biol. 61:545-550.) Chemokines were initially identified as chemotactic proteins that recruit monocytes and macrophages to sites of inflammation. Recent evidence indicates that chemokines may also play key roles in hematopoiesis and HIV-1 infection. Chemokines are small proteins which range from about 6-15 kilodaltons in molecular weight. Chemokines are further classified as C, CC, CXC, or CX_3C based on the number and position of critical cysteine residues. The CC chemokines, for example, each contain a conserved motif consisting of two consecutive cysteines followed by two additional cysteines which occur downstream at 24- and 16-residue intervals, respectively (ExPASy PROSITE database, documents PS00472 and PDOC00434). The presence and spacing of these four cysteine residues are highly conserved, whereas the intervening residues diverge significantly. However, a conserved tyrosine located about 15 residues downstream of the cysteine doublet seems to be important for chemotactic activity. Most of the human genes encoding CC chemokines are clustered on chromosome 17, although there are a few examples of CC chemokine genes that map elsewhere.

Recently, a novel CC chemokine has been identified in mouse and human thymus (Vicari, A.P. et al. (1997) Immunity 7:291-301). This protein, called thymus-expressed chemokine (TECK), is also expressed at lower levels in the small intestine. TECK likely plays a role in T-lymphocyte

development for two reasons. First, TECK is most abundantly expressed in the thymus, which is the major lymphoid organ where T-lymphocyte maturation occurs. Second, the primary source of TECK in the thymus is dendritic cells, which are leukocytic cells that help establish self-tolerance in developing T-lymphocytes. In addition, TECK demonstrates chemotactic activity for activated macrophages, dendritic cells, and thymic T-lymphocytes. The cDNA encoding human TECK (hTECK) contains an open reading frame of 453 base pairs which predicts a protein of 151 amino acids. hTECK retains the conserved features of CC chemokines described above, including four conserved cysteines at C30, C31, C58, and C75. However, the spacing between C31 and C58 is increased by three residues, and the spacing between C58 and C75 is increased by one residue. In addition, hTECK lacks the conserved tyrosine found in most CC chemokines.

The third class of growth factors includes the small peptide factors, which primarily function as hormones in the regulation of highly specialized processes other than cellular proliferation. These factors, which are typically less than 20 amino acids in length, are generated by the proteolytic processing of larger precursor proteins. Some of these factors include bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, bradykinin, and related peptides. (See, e.g., Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor MI; McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY; and Habenicht, A., ed. (1990) Growth Factors, Differentiation Factors, and Cytokines, Springer-Verlag, New York NY.)

The discovery of new human growth factor related molecules and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of developmental disorders; cell proliferative disorders including cancer; immune disorders including inflammation; reproductive and cardiovascular disorders; and infections.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human growth factor related molecules, referred to collectively as "GFRP" and individually as "GFRP-1," "GFRP-2," "GFRP-3," and "GFRP-4." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof. The invention also includes a polypeptide comprising an amino acid sequence that differs by one or more conservative amino acid substitutions from an amino acid sequence selected from the group consisting of SEQ ID NO:1-4.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID

NO:1-4 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8 and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-4. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing a polynucleotide of the invention under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-4 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of GFRP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of GFRP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

Figure 1 shows the amino acid sequence alignment between GFRP-1 (Incyte Clone 2777282; SEQ ID NO:1) and ANUP (GI 1536902; SEQ ID NO:9), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figure 2 shows the amino acid sequence alignment between GFRP-2 (Incyte Clone 4185824; SEQ ID NO:2) and hTECK (GI 2388627; SEQ ID NO:10), produced using the multisequence alignment program of LASERGENE software (DNASTAR).

Figures 3A and 3B show the amino acid sequence alignment between GFRP-3 (Incyte Clone 2484440; SEQ ID NO:3) and chicken follistatin (GI 853834; SEQ ID NO:11), produced using the multisequence alignment program of LASERGENE software (DNASTAR).

Figure 4 shows the partial amino acid sequence alignment between GFRP-4 (Incyte Clone 4163378; SEQ ID NO:4) and human bone morphogenetic protein 1, BMP-1 (GI 179500; SEQ ID NO:12), produced using the multisequence alignment program of LASERGENE software (DNASTAR).

Table 1 shows the tools, programs, and algorithms used to analyze GFRP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will

be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"GFRP" refers to the amino acid sequences of substantially purified GFRP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GFRP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GFRP either by directly interacting with GFRP or by acting on components of the biological pathway in which GFRP participates.

An "allelic variant" is an alternative form of the gene encoding GFRP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding GFRP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GFRP or a polypeptide with at least one functional characteristic of GFRP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GFRP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding

GFRP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GFRP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GFRP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GFRP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GFRP either by directly interacting with GFRP or by acting on components of the biological pathway in which GFRP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GFRP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies

which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic GFRP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding GFRP or fragments of GFRP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison

WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of GFRP or the polynucleotide encoding GFRP which is

identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:5-8 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:5-8, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:5-8 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:5-8 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:5-8 and the region of SEQ ID NO:5-8 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-4 is encoded by a fragment of SEQ ID NO:5-8. A fragment of SEQ ID NO:1-4 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-4. For example, a fragment of SEQ ID NO:1-4 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-4. The precise length of a fragment of SEQ ID NO:1-4 and the region of SEQ ID NO:1-4 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be

tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of GFRP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GFRP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably

linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding GFRP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of

Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding GFRP, or fragments thereof, or GFRP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell: a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate: a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular

structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to

each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human growth factor related molecules (GFRP), the polynucleotides encoding GFRP, and the use of these compositions for the diagnosis, treatment, or prevention of developmental disorders; cell proliferative disorders including cancer; immune disorders including inflammation; reproductive and cardiovascular disorders; and infections.

Nucleic acids encoding the GFRP-1 of the present invention were identified in Incyte Clone 2777282H1 from the ovarian tumor cDNA library (OVARTUT03) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:5, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2777282H1 (OVARTUT03) and 2777282T6 (OVARTUT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. GFRP-1 is 125 amino acids in length and has one potential N-glycosylation site at N46; two potential casein kinase II phosphorylation sites at T44 and S80; and one potential protein kinase C phosphorylation site at S122. MOTIFS analysis indicates that GFRP-1 contains a Ly-6/u-PAR domain signature from R24 to C73. Likewise, BLOCKS analysis indicates that all three protein blocks which are characteristic of Ly-6/uPAR domains are found in GFRP-1 from L9 to C28, from Q87 to N100, and from P105 to L118. SPSCAN and HMM analyses indicate that GFRP-1 contains a signal peptide sequence from M1 to either L19 or A22. As shown in Figure 1, GFRP-1 has chemical and structural similarity with ANUP (also known as ARS; GI 1536902; SEQ ID NO:9). GFRP-1 and ANUP share 39% identity and, as shown in bold type, all eleven cysteine residues in ANUP are conserved in GFRP-1. GFRP-1 also contains an additional C-terminal cysteine at C112. GFRP-1 and ANUP are both acidic proteins with predicted isoelectric points of 5.8 and 5.2, respectively. Furthermore, GFRP-1 and ANUP are both relatively small proteins of 125 and 103 amino acids in length, respectively. The Ly-6/uPAR domain signature is also conserved between the

two proteins. GFRP-1 also has chemical and structural similarity with mouse ARS component B precursor (GI 4218459). Fragments of SEQ ID NO:5 from about nucleotide 168 to about nucleotide 197 and from about nucleotide 390 to about nucleotide 419 are useful in hybridization or amplification technologies to identify SEQ ID NO:5 and to distinguish between SEQ ID NO:5 and a related sequence. Northern analysis shows the expression of this sequence in cDNA libraries derived from ovarian tumor tissue and thymus.

Nucleic acids encoding the GFRP-2 of the present invention were identified in Incyte Clone 4185824H1 from the breast tissue cDNA library (BRSTNOT31) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:6, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 4185824H1 (BRSTNOT31) and 4185824F6 (BRSTNOT31).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2. GFRP-2 is 127 amino acids in length and has one potential N-glycosylation site at N78; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S110; and three potential protein kinase C phosphorylation sites at S39, T80, and S110. SPSCAN and HMM analyses indicate that GFRP-2 contains a signal peptide sequence from M1 to either A19 or A22. Analysis of GFRP-2 suggests that this protein is a CC chemokine. The predicted molecular weight of GFRP-2 is 14.3 kilodaltons, which is typical of chemokines. The amino acid sequence of GFRP-2 from C30 to V74 shows strong similarity to the CC chemokine consensus sequence (ExPASy PROSITE database, documents PS00472, PDOC00434). The four cysteines which are characteristic of this consensus sequence are conserved at C30, C31, C58, and C73. All but one of the remaining amino acids in this region match the consensus sequence. The spacing between C31 and C58 and between C58 and C73 is slightly altered from the consensus sequence. However, similar alterations in spacing have been observed in other CC chemokines such as hTECK. As shown in Figure 2, GFRP-2 has chemical and structural similarity with hTECK (GI 2388627; SEQ ID NO:10). In particular, GFRP-2 and hTECK share 20% identity, including the four cysteines which are characteristic of the CC chemokine consensus sequence (shown in bold type). In addition, GFRP-2 and hTECK are both basic proteins with predicted isoelectric points of 10.1 and 10.2, respectively. GFRP-2 also has chemical and structural similarity with human Dvic-1 C-C chemokine (GeneSeq ID W60649) and with mouse CC chemokine ALP (GI 4140686). A fragment of SEQ ID NO:6 from about nucleotide 287 to about nucleotide 316 is useful in hybridization or amplification technologies to identify SEQ ID NO:6 and to distinguish between SEQ ID NO:6 and a related sequence. Northern analysis shows the expression of this sequence in five cDNA libraries, four of which are derived from normal or breast tumor tissue.

Nucleic acids encoding the GFRP-3 of the present invention were identified in Incyte Clone

2484440 from the smooth muscle cell cDNA library (SMCANOT01) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:7, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2484440H1 (SMCANOT01), 4763347H1 (PLACNOT05), 961852R1 (BRSTTUT03), 2026240X20C1(KERANOT02), 867727R1 (BRAITUT03), 1340443F1 (COLNTUT03), and 1513127T1 (PANCTUT01) and 1513127F1 (PANCTUT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:3. GFRP-3 is 147 amino acids in length and has a potential N-glycosylation site at residue N99, two potential casein kinase II phosphorylation sites at residues T130 and S139, and seven potential N-myristoylation sites at residues G4, G7, G8, G10, G24, G27, and G31. BLOCKS analysis indicates that the regions of GFRP-3 from residue E53 to P85 and from residue G91 to S126 share homology with an osteonectin domain. PROFILESCAN analysis likewise indicates that the region of GFRP-3 from residue Q64 to C106 shares homology with an osteonectin domain. PFAM analysis indicates that the region of GFRP-3 from residue C76 to C113 shares homology with a Kazal-type serine protease inhibitor domain. The regions identified by BLOCKS, PROFILESCAN, and PFAM comprise a domain with similarity to the polycysteine follistatin domain. GFRP-3 contains a predicted signal peptide from residue M1 to G31. As shown in Figures 3A and 3B, GFRP-3 has chemical and structural similarity with chicken follistatin (GI 853834; SEQ ID NO:11). In particular, GFRP-3 and chicken follistatin share 33% identity. In addition, the region of GFRP-3 from C55 to G120 shares 61% identity with the region of follistatin from C167 to G231. This region contains nine of the ten cysteine residues characteristic of the follistatin domain, with nearly identical spacing patterns. GFRP-3 also has chemical and structural similarity with human follistatin-related protein FLRG (GI 3764055). A fragment of SEQ ID NO:7 from about nucleotide 110 to about nucleotide 154 is useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:7 and to distinguish between SEQ ID NO:7 and a related sequence. The encoded polypeptide is useful, for example, as an immunogenic peptide. Northern analysis shows the expression of this sequence in various libraries, at least 75% of which are associated with cell proliferation, and at least 23% are associated with inflammation. Of particular note, at least 29% of the libraries expressing GFRP-3 are associated with reproductive tissue.

Nucleic acids encoding the GFRP-4 of the present invention were identified in Incyte Clone 4163378 from the diseased breast tissue cDNA library (BRSTNOT32) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:8, was assembled from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 4163378H1 and 4163378F6 (BRSTNOT32), 3598059F6 (FIBPNOT01), 2962366H1 (ADRENOT09), 3325672H2 (PTHYNOT03), 3073703H1 (BONEUNT01), 1302516F1 (PLACNOT02), 2739211F6

(OVARNOT09). 877279T1 (LUNGAST01). and 3094133F6 (BRSTNOT19).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:4. GFRP-4 is 345 amino acids in length and has three potential N-glycosylation sites at residues N25, N55, and N254; seven potential casein kinase II phosphorylation sites at T20, S34, T89, S194, T195, S258, and S323; and six potential protein kinase C phosphorylation sites at S27, S34, S60, T251, S258, and T302. GFRP-4 also contains a potential signal peptide between residues M1 and A14, as determined by HMM analysis; a PDGF family signature between residues F229 and Q310, as determined by ProfileScan; and a CUB domain between residues R48 and Y160 and a PDGF domain between residues I269 and C337, both determined by PFAM analysis. As shown in Figure 4, GFRP-4 has chemical and structural similarity with human bone morphogenetic protein 1, BMP-1 (GI 179500; SEQ ID NO:12). In particular, GFRP-4 and BMP-1 share 27% identity near the C-terminus of BMP-1 between residues 599 and 718 of BMP-1. This region of GFRP-4 encompasses the CUB domain identified above between residues 48 to 160 and includes two cysteine residues at C104 and C124, also shared by BMP-1, that are proposed to be involved in cysteine-cysteine disulfide bridging. GFRP-4 also has chemical and structural similarity with chicken bone morphogenetic protein 1 (GI 2852121). A fragment of SEQ ID NO:8 from about nucleotide 420 to about nucleotide 467 is useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:8 and to distinguish between SEQ ID NO:8 and a related sequence. The encoded polypeptide is useful, for example, as an immunogenic peptide. Northern analysis shows the expression of this sequence in various libraries, at least 57% of which are associated with cancer and at least 30% of which are associated with the inflammation and the immune response. Of particular note is the expression of GFRP-4 in reproductive and cardiovascular tissue.

The invention also encompasses GFRP variants. A preferred GFRP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GFRP amino acid sequence, and which contains at least one functional or structural characteristic of GFRP.

The invention also encompasses polynucleotides which encode GFRP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:5-8, which encodes GFRP.

The invention also encompasses a variant of a polynucleotide sequence encoding GFRP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GFRP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:5-

8 which has at least about 70%. or alternatively at least about 85%. or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:5-8. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GFRP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GFRP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus. the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GFRP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GFRP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GFRP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GFRP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GFRP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GFRP and GFRP derivatives. or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GFRP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences. and, in particular, to those shown in SEQ ID NO:5-8 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment

of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding GFRP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GFRP may be cloned in recombinant DNA molecules that direct expression of GFRP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GFRP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GFRP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding GFRP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, GFRP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of GFRP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active GFRP, the nucleotide sequences encoding GFRP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GFRP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GFRP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GFRP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GFRP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GFRP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GFRP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GFRP can be achieved using a

multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GFRP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GFRP are needed, e.g. for the production of antibodies, vectors which direct high level expression of GFRP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GFRP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of GFRP. Transcription of sequences encoding GFRP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GFRP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GFRP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are

constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of GFRP in cell lines is preferred. For example, sequences encoding GFRP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GFRP is inserted within a marker gene sequence, transformed cells containing sequences encoding GFRP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GFRP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GFRP and that express

GFRP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GFRP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GFRP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GFRP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GFRP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GFRP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GFRP may be designed to contain signal sequences which direct secretion of GFRP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation,

phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GFRP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GFRP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GFRP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GFRP encoding sequence and the heterologous protein sequence, so that GFRP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GFRP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

Fragments of GFRP may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of GFRP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GFRP and human growth factor related molecules. In particular, chemical and structural similarity exists between regions of GFRP-1 and ANUP, between regions of GFRP-2 and CC chemokines, between regions of GFRP-3 and chicken follistatin, and between GFRP-4 and growth factor related molecule. In addition, the expression of GFRP-1 is closely associated with ovarian tumor and thymic tissue; the expression of GFRP-2 is closely associated with tumorous and nontumorous breast tissue; the expression of GFRP-3 is closely associated with cell proliferation and inflammation, and with reproductive tissue; and the expression of GFRP-4 is closely associated with cancer, inflammation and the immune response, and with reproductive and cardiovascular tissue. Therefore, GFRP appears to play a role in developmental disorders; cell proliferative disorders including cancer; immune disorders including inflammation; reproductive and cardiovascular disorders; and infections. In the treatment of disorders associated with increased GFRP expression or activity, it is desirable to decrease the expression or activity of GFRP. In the treatment of disorders associated with decreased GFRP expression or activity, it is desirable to increase the expression or activity of GFRP.

Therefore, in one embodiment, GFRP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GFRP. Examples of such disorders include, but are not limited to, a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune disorder, such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis,

bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a reproductive disorder, such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a cardiovascular disorder, such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an infection such as that caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection such as that caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection such as that caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, or other fungal agents causing various mycoses; and an infection such as that caused by a parasite

classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, or cestrodes such as tapeworm.

In another embodiment, a vector capable of expressing GFRP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GFRP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified GFRP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GFRP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of GFRP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GFRP including, but not limited to, those listed above.

In a further embodiment, an antagonist of GFRP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GFRP. Examples of such disorders include, but are not limited to, those developmental disorders; cell proliferative disorders including cancer; immune disorders including inflammation; reproductive and cardiovascular disorders; and infections described above. In one aspect, an antibody which specifically binds GFRP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GFRP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GFRP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GFRP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of GFRP may be produced using methods which are generally known in the art. In particular, purified GFRP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GFRP. Antibodies to GFRP may also be generated using methods that are well known in the art. Such antibodies may include, but are not

limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GFRP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GFRP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of GFRP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to GFRP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GFRP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA

86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for GFRP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GFRP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering GFRP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GFRP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of GFRP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GFRP epitopes, represents the average affinity, or avidity, of the antibodies for GFRP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GFRP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the GFRP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GFRP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies. Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GFRP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty,

supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding GFRP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding GFRP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding GFRP. Thus, complementary molecules or fragments may be used to modulate GFRP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GFRP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding GFRP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding GFRP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding GFRP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding GFRP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GFRP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding GFRP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the

therapeutic effects discussed above. Such pharmaceutical compositions may consist of GFRP, antibodies to GFRP, and mimetics, agonists, antagonists, or inhibitors of GFRP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol.

Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of GFRP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for

administration in humans.

A therapeutically effective dose refers to that amount of active ingredient. for example GFRP or fragments thereof, antibodies of GFRP, and agonists, antagonists or inhibitors of GFRP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GFRP may be used for the diagnosis of disorders characterized by expression of GFRP, or in assays to monitor patients being treated with GFRP or agonists, antagonists, or inhibitors of GFRP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GFRP include methods which utilize the antibody and a label to detect GFRP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of

reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring GFRP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GFRP expression. Normal or standard values for GFRP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to GFRP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GFRP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GFRP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GFRP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GFRP, and to monitor regulation of GFRP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GFRP or closely related molecules may be used to identify nucleic acid sequences which encode GFRP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding GFRP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the GFRP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:5-8 or from genomic sequences including promoters, enhancers, and introns of the GFRP gene.

Means for producing specific hybridization probes for DNAs encoding GFRP include the cloning of polynucleotide sequences encoding GFRP or GFRP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding GFRP may be used for the diagnosis of disorders associated with expression of GFRP. Examples of such disorders include, but are not limited to, a

developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune disorder, such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a reproductive disorder, such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a

cardiovascular disorder, such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an infection such as that caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection such as that caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection such as that caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, or other fungal agents causing various mycoses; and an infection such as that caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, or cestrodes such as tapeworm. The polynucleotide sequences encoding GFRP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered GFRP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GFRP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GFRP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GFRP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GFRP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GFRP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GFRP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding GFRP, or a fragment of a polynucleotide complementary to the polynucleotide encoding GFRP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of GFRP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the

polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding GFRP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GFRP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be

used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GFRP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GFRP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GFRP, or fragments thereof, and washed. Bound GFRP is then detected by methods well known in the art. Purified GFRP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GFRP specifically compete with a test compound for binding GFRP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GFRP.

In additional embodiments, the nucleotide sequences which encode GFRP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0627 P, filed October 28, 1998], U.S. Ser. No. [Attorney Docket No. PF-0644 P, filed December 11, 1998], and U.S. Ser. No. [Attorney Docket No. PF-0688 P, filed May 17, 1999], are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

The OVARTUT03 cDNA library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 52-year-old mixed ethnicity female during a total abdominal hysterectomy, bilateral salpingo-oophorectomy, peritoneal and lymphatic structure biopsy, regional lymph node excision, and peritoneal tissue destruction. Pathology indicated an invasive grade 3 (of 4) seroanaplastic carcinoma forming a mass in the left ovary. Multiple tumor implants were present on the surface of the ovaries and fallopian tubes, the posterior surface of the uterus, and cul-de-sac. Multiple leiomyomata were identified. Pathology also indicated metastatic grade 3 seroanaplastic carcinoma involving the omentum, cul-de-sac peritoneum, left broad ligament peritoneum, and mesentery colon. Patient history included breast cancer, chronic peptic ulcer, and joint pain. Family history included colon cancer, cerebrovascular disease, breast cancer, type II diabetes, esophageal cancer, and depressive disorder.

The BRSTNOT31 cDNA library was constructed using RNA isolated from right breast tissue removed from a 57-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated residual microscopic infiltrating grade 3 ductal adenocarcinoma and extensive grade 2 intraductal carcinoma. Multiple axillary lymph nodes were positive for metastatic adenocarcinoma with minimal extranodal extension. Immunoperoxidase stains for estrogen and progesterone receptors were positive. Patient history included benign hypertension, hyperlipidemia, cardiac dysrhythmia, benign colon neoplasm, breast cyst, and breast neoplasm. Family history included benign hypertension, acute leukemia, primary liver cancer, and lung cancer.

The SMCANOT01 library was constructed using RNA isolated from an aortic smooth muscle cell line derived from the explanted heart of a male during a heart transplant.

The BRSTNOT32 library was constructed from RNA isolated from diseased right breast tissue removed from a 46-year-old Caucasian female during bilateral subcutaneous mammectomy. Pathology indicated nonproliferative fibrocystic disease bilaterally. The patient presented with fibrosclerosis of the breast. Patient history included urinary tract infection. Family history included breast cancer, benign hypertension, and atherosclerotic coronary artery disease.

For the OVARTUT03 and BRSTNOT31 libraries, frozen tissue was homogenized and lysed in TRIZOL reagent (1 g tissue/10 ml TRIZOL; Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate, using a Polytron PT-3000 homogenizer (Brinkmann Instruments, Westbury NY). After brief incubation on ice, chloroform was added (1:5 v/v), and the mixture was centrifuged to separate the phases. The upper aqueous phase was removed to a fresh tube, and isopropanol was added to precipitate RNA. The RNA was resuspended in RNase-free water and

treated with DNase. The RNA was re-extracted once with acid phenol-chloroform and reprecipitated with sodium acetate and ethanol.

For the SMCANOT01 library, the frozen tissue was homogenized and lysed in guanidinium isothiocyanate solution using a Polytron-PT 3000 homogenizer (Brinkmann Instruments). RNA was isolated as per Stratagene's RNA isolation protocol (Stratagene, La Jolla CA). RNA was extracted twice with acid phenol, precipitated with sodium acetate and ethanol, resuspended in RNase-free water, and treated with DNase.

For the BRSTNOT32 library, the frozen tissue was homogenized and lysed in Trizol reagent (0.8 g tissue/12 ml Trizol; Life Technologies), a monoplastic solution of phenol and guanidine isothiocyanate, using a Polytron PT-3000 homogenizer (Brinkmann Instruments). After a brief incubation on ice, chloroform was added (1:5 v/v) and the lysate was centrifuged. The upper chloroform layer was removed to a fresh tube and the RNA precipitated with isopropanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37°C. The mRNA was re-extracted once with acid phenol-chloroform pH 4.7 and precipitated using 0.3M sodium acetate and 2.5 volumes ethanol.

To construct the OVARTUT03, BRSTNOT31, SMCANOT01, and BRSTNOT32 cDNA libraries, mRNA was isolated using the QIAGEN OLIGOTEX kit (QIAGEN, Inc., Chatsworth CA). The mRNA was handled according to the recommended protocols in the SUPERSCRIPT plasmid system for cDNA synthesis and plasmid cloning (Life Technologies). The cDNAs were fractionated on a SEPHAROSE CL4B column (Pharmacia), and those cDNAs exceeding 400 bp were ligated into either the pINCY or pINCY1 plasmids. The recombinant plasmids were subsequently transformed into DH5α competent cells (Life Technologies).

II. Isolation of cDNA Clones

Plasmid DNA was released from the cells and purified using the R.E.A.L. PREP 96 plasmid kit (QIAGEN, Inc.). This kit enabled the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Life Technologies) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a

high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a Fluoroskan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 1 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 1 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:5-8. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are

usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding GFRP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in the description of the invention.

V. Extension of GFRP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:5-8 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN

quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:5-8 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:5-8 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments.

Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer. 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech). and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus. Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by

procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the GFRP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GFRP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GFRP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GFRP-encoding transcript.

IX. Expression of GFRP

Expression and purification of GFRP is achieved using bacterial or virus-based expression systems. For expression of GFRP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express GFRP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GFRP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GFRP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, GFRP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham

Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GFRP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GFRP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of GFRP Activity

An assay for cytokine activity of GFRP measures the proliferation of leukocytes. In this assay, the amount of tritiated thymidine incorporated into newly synthesized DNA is used to estimate proliferative activity. Varying amounts of GFRP are added to cultured leukocytes, such as granulocytes, monocytes, or lymphocytes, in the presence of [^{3}H]thymidine, a radioactive DNA precursor. A similar assay for growth factor activity of GFRP measures the stimulation of DNA synthesis in Swiss mouse 3T3 cells (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY). Initiation of DNA synthesis indicates the cells' entry into the mitotic cycle and their commitment to undergo later division. 3T3 cells are competent to respond to most growth factors, not only those that are mitogenic, but also those that are involved in embryonic induction. This competency is possible because the in vivo specificity demonstrated by some growth factors is not necessarily inherent but is determined by the responding tissue. Therefore, this assay is generally applicable to GFRP. In this growth factor assay, varying amounts of GFRP are added to quiescent 3T3 cultured cells in the presence of [^{3}H]thymidine. GFRP for these cytokine and growth factor assays can be obtained by recombinant means or from biochemical preparations. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GFRP concentration range is indicative of cytokine or growth factor activity, as appropriate. One unit of activity per milliliter is conventionally defined as the concentration of GFRP producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA.

An assay for chemokine activity of GFRP utilizes a Boyden micro chamber (Neuroprobe, Cabin John MD) to measure leukocyte chemotaxis (Vicari, supra). In this assay, about 10^5 migratory cells such as macrophages or monocytes are placed in cell culture media in the upper compartment of the chamber. Varying dilutions of GFRP are placed in the lower compartment. The two compartments are separated by a 5 or 8 micron pore polycarbonate filter (Nucleopore, Pleasanton

CA). After incubation at 37°C for 80 to 120 minutes, the filters are fixed in methanol and stained with appropriate labeling agents. Cells which migrate to the other side of the filter are counted using standard microscopy. The chemotactic index is calculated by dividing the number of migratory cells counted when GFRP is present in the lower compartment by the number of migratory cells counted when only media is present in the lower compartment. The chemotactic index is proportional to the activity of GFRP.

The follistatin activity of GFRP is measured by its binding to activin using a sensitive and specific assay for activin-follistatin binding developed by Demura et al. (1992, Biochem. Biophys. Res. Commun. 185:1148-1154). Activin is labelled with ^{125}I and combined with a biological sample. The mixture is treated with 50% acetonitrile in conjunction with chromatographic methods which separate free from bound activin. The amount of GFRP bound to activin is directly proportional to its follistatin activity.

XI. Functional Assays

GFRP function is assessed by expressing the sequences encoding GFRP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are

discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GFRP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GFRP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GFRP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of GFRP Specific Antibodies

GFRP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GFRP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-GFRP activity by, for example, binding the peptide or GFRP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring GFRP Using Specific Antibodies

Naturally occurring or recombinant GFRP is substantially purified by immunoaffinity chromatography using antibodies specific for GFRP. An immunoaffinity column is constructed by covalently coupling anti-GFRP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GFRP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GFRP (e.g., high ionic strength

buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/GFRP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GFRP is collected.

XIV. Identification of Molecules Which Interact with GFRP

GFRP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GFRP, washed, and any wells with labeled GFRP complex are assayed. Data obtained using different concentrations of GFRP are used to calculate values for the number, affinity, and association of GFRP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 1 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:5-8 and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.

12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

18. A purified antagonist of the polypeptide of claim 1.

19. A method for treating or preventing a disorder associated with decreased expression or activity of GFRP, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a disorder associated with increased expression or activity of GFRP, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

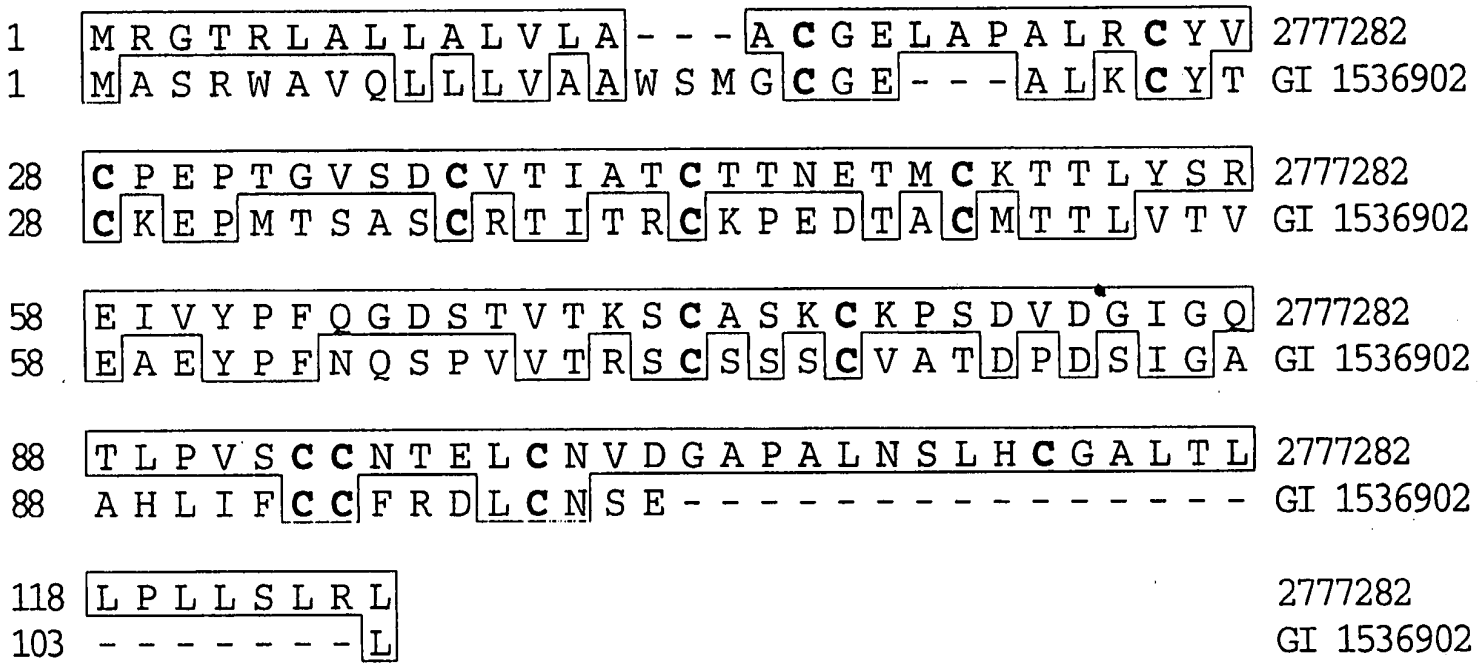
1 MRGTRLALLALVLA---ACGELAPALRCYV 2777282
1 MASRWAVQLLLVAAWSMGCGE---ALKCYT GI 1536902
28 CPEPTGVSDCVTIATCTTNETMCKTTLYSR 2777282
28 CKEPMTSASCRTITRCKPEDTACMTTLVTV GI 1536902
58 EIVYPFQGDSTVTKSCASKCKPSDVDGIGQ 2777282
58 EAEYPFNQSPVVTRSCSSSCVATDPDSIGA GI 1536902
88 TLPVSCCNTELCNVDGAPALNSLHCGALTL 2777282
88 AHLIFCCFRDLCNSE---------------- GI 1536902
118 LPLLSLRL 2777282
103 --------L GI 1536902

FIGURE 1

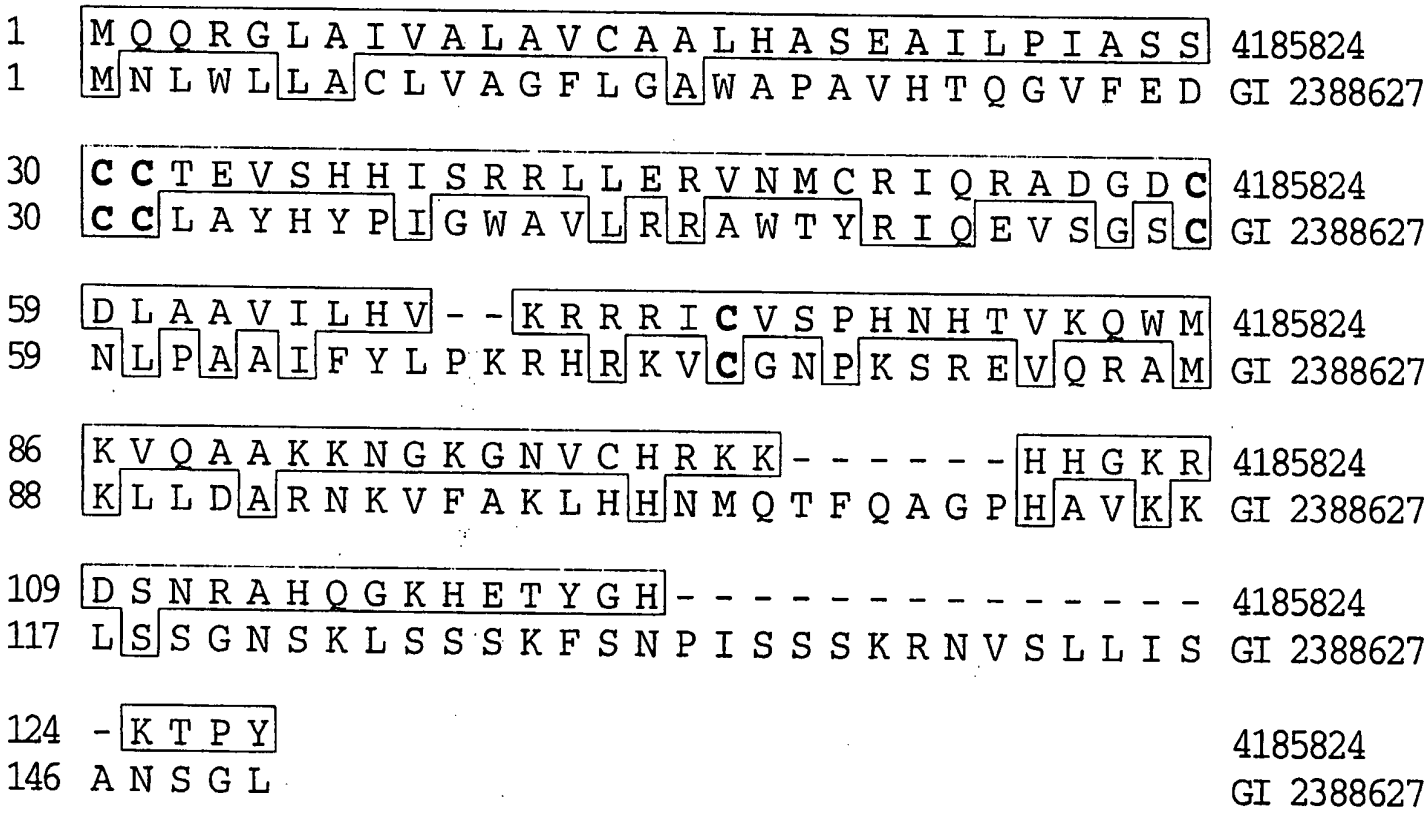
1 MQQRGLAIVALAVCAALHASEAILPIASS 4185824
1 MNLWLLACLVAGFLGAWAPAVHTQGVFED GI 2388627
30 CCTEVSHHISRRLLERVNMCRIQRADGDC 4185824
30 CCLAYHYPIGWAVLRRAWTYRIQEVSGSC GI 2388627
59 DLAAVILHV--KRRRICVSPHNHTVKQWM 4185824
59 NLPAAIFYLPKRHRKVCGNPKSREVQRAM GI 2388627
86 KVQAAKKNGKGNVCHRKK------HHGKR 4185824
88 KLLDARNKVFAKLHHNMQTFQAGPHAVKK GI 2388627
109 DSNRAHQGKHETYGH---------------- 4185824
117 LSSGNSKLSSSKFSNPISSSKRNVSLLIS GI 2388627
124 -KTPY 4185824
146 ANSGL GI 2388627

FIGURE 2

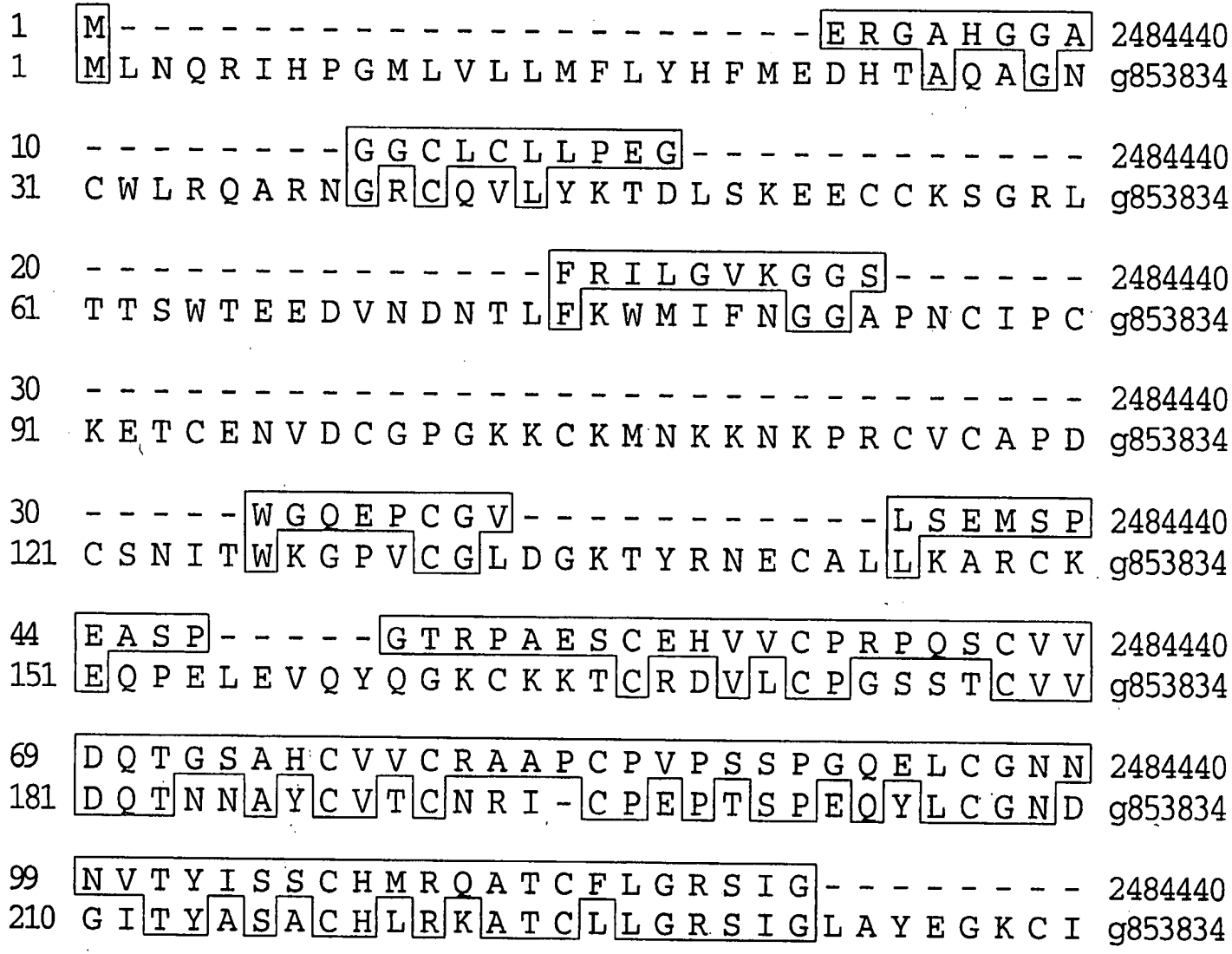
1 M - E R G A H G G A 2484440
1 M L N Q R I H P G M L V L L M F L Y H F M E D H T A Q A G N g853834
10 - - - - - - - - - G G C L C L L P E G - - - - - - - - - - - - - 2484440
31 C W L R Q A R N G R C Q V L Y K T D L S K E E C C K S G R L g853834
20 - - - - - - - - - - - - - - - - F R I L G V K G G S - - - - - - 2484440
61 T T S W T E E D V N D N T L F K W M I F N G G A P N C I P C g853834
30 - 2484440
91 K E T C E N V D C G P G K K C K M N K K N K P R C V C A P D g853834
30 - - - - - W G Q E P C G V - - - - - - - - - - - - L S E M S P 2484440
121 C S N I T W K G P V C G L D G K T Y R N E C A L L K A R C K g853834
44 E A S P - - - - - G T R P A E S C E H V V C P R P Q S C V V 2484440
151 E Q P E L E V Q Y Q G K C K K T C R D V L C P G S S T C V V g853834
69 D Q T G S A H C V V C R A A P C P V P S S P G Q E L C G N N 2484440
181 D Q T N N A Y C V T C N R I - C P E P T S P E Q Y L C G N D g853834
99 N V T Y I S S C H M R Q A T C F L G R S I G - - - - - - - - - 2484440
210 G I T Y A S A C H L R K A T C L L G R S I G L A Y E G K C I g853834

FIGURE 3A

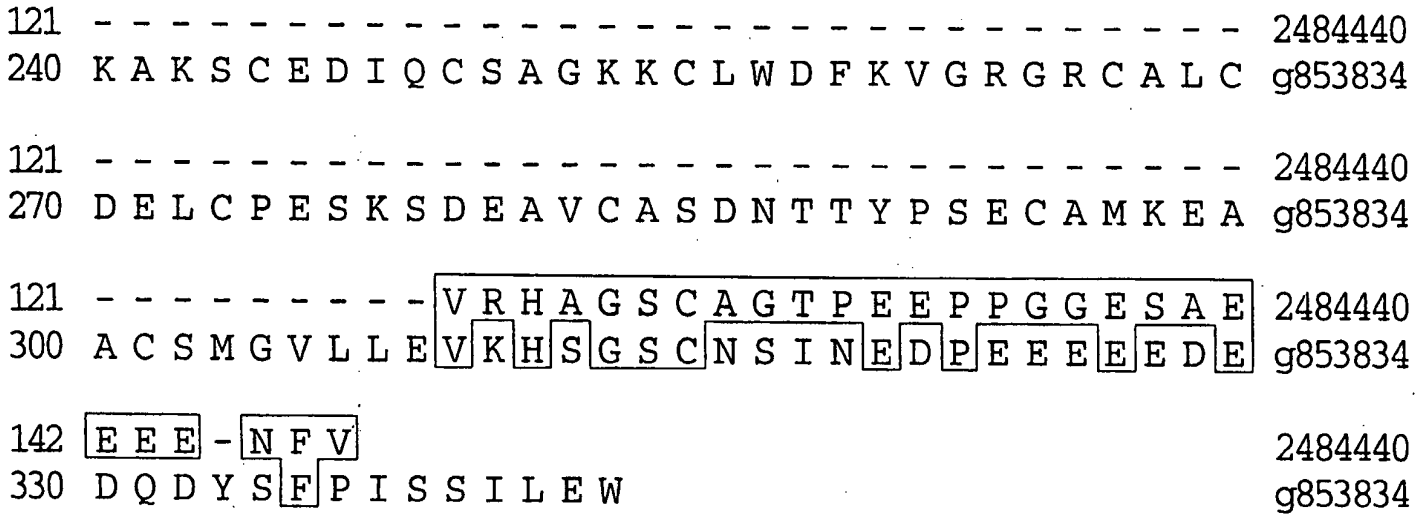
121 - 2484440
240 K A K S C E D I Q C S A G K K C L W D F K V G R G R C A L C g853834
121 - 2484440
270 D E L C P E S K S D E A V C A S D N T T Y P S E C A M K E A g853834
121 - - - - - - - - - - V R H A G S C A G T P E E P P G G E S A E 2484440
300 A C S M G V L L E V K H S G S C N S I N E D P E E E E E D E g853834
142 E E E - N F V 2484440
330 D Q D Y S F P I S S I L E W g853834

FIGURE 3B

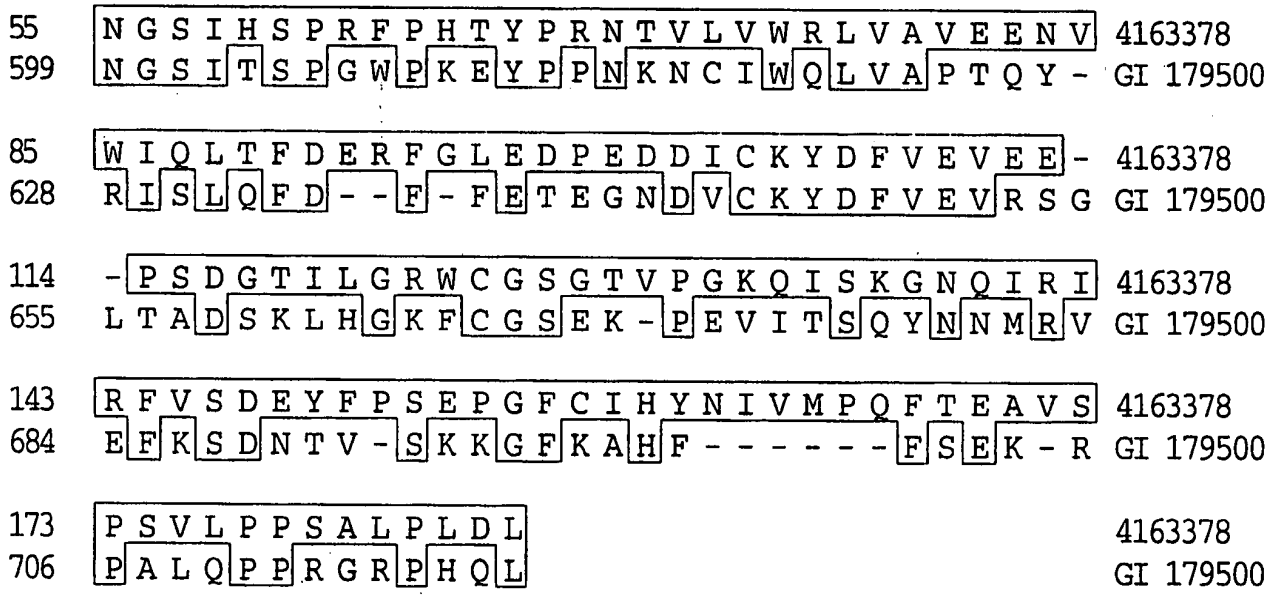
55 NGSIHSPRFPHTYPRNTVLVWRLVAVEENV 4163378
599 NGSITSPGWPKEYPPNKNCIWQLVAPTQY- GI 179500
85 WIQLTFDERFGLEDPEDDICKYDFVEVEE- 4163378
628 RISLQFD--F-FETEGNDVCKYDFVEVRSG GI 179500
114 -PSDGTILGRWCGSGTVPGKQISKGNQIRI 4163378
655 LTADSKLHGKFCGSEK-PEVITSQYNNMRV GI 179500
143 RFVSDEYFPSEPGFCIHYNIVMPQFTEAVS 4163378
684 EFKSDNTV-SKKGFKAHF------FSEK-R GI 179500
173 PSVLPPSALPLDL 4163378
706 PALQPPRGRPHQL GI 179500

FIGURE 4

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
TANG, Y. Tom
YUE, Henry
HILLMAN, Jennifer L.
CORLEY, Neil C.
GUEGLER, Karl J.
BAUGHN, Mariah R.
AU-YOUNG, Janice

<120> GROWTH FACTOR RELATED MOLECULES

<130> PF-0627 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/181,711; unassigned; 09/209,547; unassigned; 09/313,457; unassigned
<151> 1998-10-28; 1998-10-28; 1998-12-11; 1998-12-11; 1999-05-17; 1999-05-17

<160> 12

<170> PERL Program

<210> 1
<211> 125
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2777282CD1

<400> 1

Met	Arg	Gly	Thr	Arg	Leu	Ala	Leu	Leu	Ala	Leu	Val	Leu	Ala	Ala	
1				5					10						15
Cys	Gly	Glu	Leu	Ala	Pro	Ala	Leu	Arg	Cys	Tyr	Val	Cys	Pro	Glu	
			20					25							30
Pro	Thr	Gly	Val	Ser	Asp	Cys	Val	Thr	Ile	Ala	Thr	Cys	Thr	Thr	
			35					40							45
Asn	Glu	Thr	Met	Cys	Lys	Thr	Thr	Leu	Tyr	Ser	Arg	Glu	Ile	Val	
			50					55							60
Tyr	Pro	Phe	Gln	Gly	Asp	Ser	Thr	Val	Thr	Lys	Ser	Cys	Ala	Ser	
			65					70							75
Lys	Cys	Lys	Pro	Ser	Asp	Val	Asp	Gly	Ile	Gly	Gln	Thr	Leu	Pro	
			80					85							90
Val	Ser	Cys	Cys	Asn	Thr	Glu	Leu	Cys	Asn	Val	Asp	Gly	Ala	Pro	
			95					100							105
Ala	Leu	Asn	Ser	Leu	His	Cys	Gly	Ala	Leu	Thr	Leu	Leu	Pro	Leu	
			110					115							120
Leu	Ser	Leu	Arg	Leu											

125

<210> 2
<211> 127
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 4185824CD1

<400> 2
Met Gln Gln Arg Gly Leu Ala Ile Val Ala Leu Ala Val Cys Ala
1 5 10 15
Ala Leu His Ala Ser Glu Ala Ile Leu Pro Ile Ala Ser Ser Cys
20 25 30
Cys Thr Glu Val Ser His His Ile Ser Arg Arg Leu Leu Glu Arg
35 40 45
Val Asn Met Cys Arg Ile Gln Arg Ala Asp Gly Asp Cys Asp Leu
50 55 60
Ala Ala Val Ile Leu His Val Lys Arg Arg Arg Ile Cys Val Ser
65 70 75
Pro His Asn His Thr Val Lys Gln Trp Met Lys Val Gln Ala Ala
80 85 90
Lys Lys Asn Gly Lys Gly Asn Val Cys His Arg Lys Lys His His
95 100 105
Gly Lys Arg Asp Ser Asn Arg Ala His Gln Gly Lys His Glu Thr
110 115 120
Tyr Gly His Lys Thr Pro Tyr
125

<210> 3
<211> 147
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2484440CD1

<400> 3
Met Glu Arg Gly Ala His Gly Gly Ala Gly Gly Cys Leu Cys Leu
1 5 10 15
Leu Pro Glu Gly Phe Arg Ile Leu Gly Val Lys Gly Gly Ser Trp
20 25 30
Gly Gln Glu Pro Cys Gly Val Leu Ser Glu Met Ser Pro Glu Ala
35 40 45
Ser Pro Gly Thr Arg Pro Ala Glu Ser Cys Glu His Val Val Cys
50 55 60
Pro Arg Pro Gln Ser Cys Val Val Asp Gln Thr Gly Ser Ala His
65 70 75
Cys Val Val Cys Arg Ala Ala Pro Cys Pro Val Pro Ser Ser Pro
80 85 90

```
Gly Gln Glu Leu Cys Gly Asn Asn Asn Val Thr Tyr Ile Ser Ser
                 95                  100                 105
Cys His Met Arg Gln Ala Thr Cys Phe Leu Gly Arg Ser Ile Gly
                110                 115                 120
Val Arg His Ala Gly Ser Cys Ala Gly Thr Pro Glu Glu Pro Pro
                125                 130                 135
Gly Gly Glu Ser Ala Glu Glu Glu Glu Asn Phe Val
                140                 145
```

<210> 4
<211> 345
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 4163378CD1

<400> 4

```
Met Ser Leu Phe Gly Leu Leu Leu Leu Thr Ser Ala Leu Ala Gly
  1               5                  10                  15
Gln Arg Gln Gly Thr Gln Ala Glu Ser Asn Leu Ser Ser Lys Phe
                 20                  25                  30
Gln Phe Ser Ser Asn Lys Glu Gln Tyr Gly Val Gln Asp Pro Gln
                 35                  40                  45
His Glu Arg Ile Ile Thr Val Ser Thr Asn Gly Ser Ile His Ser
                 50                  55                  60
Pro Arg Phe Pro His Thr Tyr Pro Arg Asn Thr Val Leu Val Trp
                 65                  70                  75
Arg Leu Val Ala Val Glu Glu Asn Val Trp Ile Gln Leu Thr Phe
                 80                  85                  90
Asp Glu Arg Phe Gly Leu Glu Asp Pro Glu Asp Asp Ile Cys Lys
                 95                 100                 105
Tyr Asp Phe Val Glu Val Glu Glu Pro Ser Asp Gly Thr Ile Leu
                110                 115                 120
Gly Arg Trp Cys Gly Ser Gly Thr Val Pro Gly Lys Gln Ile Ser
                125                 130                 135
Lys Gly Asn Gln Ile Arg Ile Arg Phe Val Ser Asp Glu Tyr Phe
                140                 145                 150
Pro Ser Glu Pro Gly Phe Cys Ile His Tyr Asn Ile Val Met Pro
                155                 160                 165
Gln Phe Thr Glu Ala Val Ser Pro Ser Val Leu Pro Pro Ser Ala
                170                 175                 180
Leu Pro Leu Asp Leu Leu Asn Asn Ala Ile Thr Ala Phe Ser Thr
                185                 190                 195
Leu Glu Asp Leu Ile Arg Tyr Leu Glu Pro Glu Arg Trp Gln Leu
                200                 205                 210
Asp Leu Glu Asp Leu Tyr Arg Pro Thr Trp Gln Leu Leu Gly Lys
                215                 220                 225
Ala Phe Val Phe Gly Arg Lys Ser Arg Val Val Asp Leu Asn Leu
                230                 235                 240
Leu Thr Glu Glu Val Arg Leu Tyr Ser Cys Thr Pro Arg Asn Phe
                245                 250                 255
Ser Val Ser Ile Arg Glu Glu Leu Lys Arg Thr Asp Thr Ile Phe
```

```
                260                 265                 270
Trp Pro Gly Cys Leu Leu Val Lys Arg Cys Gly Gly Asn Cys Ala
                275                 280                 285
Cys Cys Leu His Asn Cys Asn Glu Cys Gln Cys Val Pro Ser Lys
                290                 295                 300
Val Thr Lys Lys Tyr His Glu Val Leu Gln Leu Arg Pro Lys Thr
                305                 310                 315
Gly Val Arg Gly Leu His Lys Ser Leu Thr Asp Val Ala Leu Glu
                320                 325                 330
His His Glu Glu Cys Asp Cys Val Cys Arg Gly Ser Thr Gly Gly
                335                 340                 345
```

<210> 5
<211> 525
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2777282CB1

<400> 5
ccagtctgtc gccacctcac ttggtgtctg ctgtccccgc caggcaagcc tggggtgaga 60
gcacagagga gtgggccggg accatgcggg ggacgcggct ggcgctcctg gcgctggtgc 120
tggctgcctg cggagagctg gcgccggccc tgcgctgcta cgtctgtccg gagcccacag 180
gagtgtcgga ctgtgtcacc atcgccacct gcaccaccaa cgaaaccatg tgcaagacca 240
cactctactc ccgggagata gtgtaccccct tccaggggga ctccacggtg accaagtcct 300
gtgccagcaa gtgtaagccc tcggatgtgg atggcatcgg ccagaccctg cccgtgtcct 360
gctgcaatac tgagctgtgc aatgtagacg gggcgcccgc tctgaacagc ctccactgcg 420
gggccctcac gctcctccca ctcttgagcc tccgactgta gagtccccgc ccacccccat 480
ggccctatgc ggcccagccc cgaatgcctt gaagaagtcc ccccg 525

<210> 6
<211> 566
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 4185824CB1

<400> 6
tcgaacagcc tcacttgtgt tgctgtcagt gccagtaggg caggcaggaa tgcagcagag 60
aggactcgcc atcgtggcct tggctgtctg tgcggcccta catgcctcag aagccatact 120
tcccattgcc tccagctgtt gcacggaggt ttcacatcat atttccagaa ggctcctgga 180
aagagtgaat atgtgtcgca tccagagagc tgatggggat tgtgacttgg ctgctgtcat 240
ccttcatgtc aagcgcagaa gaatctgtgt cagcccgcac aaccatactg ttaagcagtg 300
gatgaaagtg caagctgcca agaaaaatgg taaaggaaat gtttgccaca ggaagaaaca 360
ccatggcaag agggacagta acagggcaca tcaggggaaa cacgaaacat acggccataa 420
aactccttat taggagagtc taccggtaaa tccttccgag accattccct caagtggact 480
ttggccctgg attgggtgta agttttatca tcctgaattc tcccctaatg ttgggccacc 540
ggaccaaacc caaatatttg gttttt 566

<210> 7
<211> 2246
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2484440CB1

<400> 7
aggggcgggg ctttgggtag cataaggggc aggagctcac gaaagaggtg tggctcctgg 60
gtacaatggg aagggcaggg cttgtgggtg gagcctgagg ggtggaatgg agaggggggc 120
tcacgggggg gcgggggggt gcttgtgtct cctaccagag ggttttcgca tcttgggggt 180
taagggtggg tcttggggcc aagagccctg cggggttctc agcgagatgt cccctgaagc 240
ttcccctggc acccggcctg cagagtcctg tgagcacgtg gtgtgcccgc ggccacagtc 300
gtgcgtcgtg gaccagacgg gcagcgccca ctgcgtggtg tgtcgagcgg cgccctgccc 360
tgtgccctcc agccccggcc aggagctttg cggcaacaac aacgtcacct acatctcctc 420
gtgccacatg cgccaggcca cctgcttcct gggccgctcc atcggcgtgc gccacgcggg 480
cagctgcgca ggcacccctg aggagccgcc aggtggtgag tctgcagaag aggaagagaa 540
cttcgtgtga gcctgcagga caggcctggg cctggtgccc gaggcccccc atcatcccct 600
gttatttatt gccacagcag agtctaattt atatgccacg gacactcctt agagcccgga 660
ttcggaccac ttggggatcc cagaacctcc ctgacgatat cctggaagga ctgaggaagg 720
gaggcctggg ggccggctgg tgggtgggat agacctgcgt tccggacact gagcgcctga 780
tttagggccc ttctctagga tgccccagcc cctaccctaa gacctattgc cggggaggat 840
tccacacttc cgctcctttg gggataaacc tattaattat tgctactatc aagagggctg 900
ggcattctct gctggtaatt cctgaagagg catgactgct tttctcagcc ccaagcctct 960
agtctgggtg tgtacggagg gtctagcctg ggtgtgtacg gagggtctag cctgggtgag 1020
tacggagggt ctagcctggg tgagtacgga gggtctagcc tgggtgagta cggagagtct 1080
agcctgggtg tgtatggagg atctagcctg ggtgagtatg gagggtctag cctgggtgag 1140
tatggagggt ctagcctggg tgtgtatgga gggtctagcc tgggtgagta tggagggtct 1200
agcctgggtg tgtatggagg gtctagcctg ggtgagtatg gagggtctag cctgggtgtg 1260
tacggagggt ctagtctgag tgcgtgtggg gacctcagaa cactgtgacc ttagcccagc 1320
aagccaggcc cttcatgaag gccaagaagg ctgccaccat tccctgccag cccaagaact 1380
ccagcttccc cactgcctct gtgtgccccc ttgcgtcctg tgaaggccat tgagaaatgc 1440
ccagtgtgcc ccctgggaaa gggcacggcc tgtgctcctg acacgggctg tgcttggcca 1500
cagaaccacc cagcgtctcc cctgctgctg tccacgtcag ttcatgaggc aacgtcgcgt 1560
ggtctcagac gtggagcagc cagcggcagc tcagagcagg gcactgtgtc cggcggagcc 1620
aagtccactc tgggggagct ctggcgggga ccacgggcca ctgctcaccc actggccccg 1680
agggggggtgt agacgccaag actcacgcat gtgtgacatc cggagtcctg gagccgggtg 1740
tcccagtggc accactaggt gcctgctgcc tccacagtgg ggttcacacc cagggctcct 1800
tggtccccca caacctgccc cggccaggcc tgcagaccca gactccagcc agacctgcct 1860
cacccaccaa tgcagccggg gctggcgaca ccagccaggt gctggtcttg ggccagttct 1920
cccacgacgg ctcaccctcc cctccatctg cgttgatgct cagaatcgcc tacctgtgcc 1980
tgcgtgtaaa ccacagcctc agaccagcta tggggagagg acaacacgga ggatatccag 2040
cttccccggt ctggggtgag gagtgtgggg agcttgggca tcctcctcca gcctcctcca 2100
gcccccaggc agtgccttac ctgtggtgcc cagaaaagtg cccctaggtt ggtgggtcta 2160
caggagcctc agccaggcag cccaccccac cctggggccc tgcctcacca aggaaataaa 2220
gactcaaaga agccaaaaaa aaaaaa 2246

<210> 8
<211> 2779
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No.: 4163378CB1

<400> 8
atttgtttaa accttgggaa actggttcag gtccaggttt tgctttgatc cttttcaaaa 60
actggagaca cagaagaggg ctctaggaaa aagttttgga tgggattatg tggaaactac 120
cctgcgattc tctgctgcca gagcaggctc ggcgcttcca ccccagtgca gccttccccc 180
ggcggtggtg aaagagactc gggagtcgct gcttccaaag tgcccgccgt gagtgagctc 240
tcaccccagt cagccaaatg agcctcttcg ggcttctcct gctgacatct gccctggccg 300
gccagagaca ggggactcag gcggaatcca acctgagtag taaattccag ttttccagca 360
acaaggaaca gtacggagta caagatcctc agcatgagag aattattact gtgtctacta 420
atggaagtat tcacagccca aggtttcctc atacttatcc aagaaatacg gtcttggtat 480
ggagattagt agcagtagag gaaaatgtat ggatacaact tacgtttgat gaaagatttg 540
ggcttgaaga cccagaagat gacatatgca agtatgattt tgtagaagtt gaggaaccca 600
gtgatggaac tatattaggg cgctggtgtg gttctggtac tgtaccagga aaacagattt 660
ctaaaggaaa tcaaattagg ataagatttg tatctgatga atattttcct tctgaaccag 720
ggttctgcat ccactacaac attgtcatgc cacaattcac agaagctgtg agtccttcag 780
tgctaccccc ttcagctttg ccactggacc tgcttaataa tgctataact gcctttagta 840
ccttggaaga ccttattcga tatcttgaac cagagagatg gcagttggac ttagaagatc 900
tatataggcc aacttggcaa cttcttggca aggcttttgt ttttggaaga aaatccagag 960
tggtggatct gaaccttcta acagaggagg taagattata cagctgcaca cctcgtaact 1020
tctcagtgtc cataagggaa gaactaaaga gaaccgatac cattttctgg ccaggttgtc 1080
tcctggttaa acgctgtggt gggaactgtg cctgttgtct ccacaattgc aatgaatgtc 1140
aatgtgtccc aagcaaagtt actaaaaaat accacgaggt ccttcagttg agaccaaaga 1200
ccggtgtcag gggattgcac aaatcactca ccgacgtggc cctggagcac catgaggagt 1260
gtgactgtgt gtgcagaggg agcacaggag gatagccgca tcaccaccag cagctcttgc 1320
ccagagctgt gcagtgcagt ggctgattct attagagaac gtatgcgtta tctccatcct 1380
taatctcagt tgtttgcttc aaggaccttt catcttcagg atttacagtg cattctgaaa 1440
gaggagacat caaacagaat taggagttgt gcaacagctc ttttgagagg aggcctaaag 1500
gacaggagaa aaggtcttca atcgtggaaa gaaaattaaa tgttgtatta aatagatcac 1560
cagctagttt cagagttacc atgtacgtat tccactagct gggttctgta tttcagttct 1620
ttcgatacgg cttagggtaa tgtcagtaca ggaaaaaaac tgtgcaagtg agcacctgat 1680
tccgttgcct tgcttaactc taaagctcca tgtcctgggc ctaaaatcgt ataaaatctg 1740
gatttttttt ttttttttg ctcatattca catatgtaaa ccagaacatt ctatgtacta 1800
caaacctggt ttttaaaaag gaactatgtt gctatgaatt aaacttgtgt cgtgctgata 1860
ggacagactg gatttttcat atttcttatt aaaatttctg ccatttagaa gaagagaact 1920
acattcatgg tttggaagag ataaacctga aaagaagagt ggccttatct tcactttatc 1980
gataagtcag tttatttgtt tcattgtgta cattttata ttctcctttt gacattataa 2040
ctgttggctt ttctaatctt gttaaatata tctatttta ccaaaggtat ttaatattct 2100
tttttatgac aacttagatc aactattttt agcttggtaa atttttctaa acacaattgt 2160
tatagccaga ggaacaaaga tgatataaaa tattgttgct ctgacaaaaa tacatgtatt 2220
tcattctcgt atggtgctag agttagatta atctgcattt taaaaaactg aattggaata 2280
gaattggtaa gttgcaaaga ctttttgaaa ataattaaat tatcatatct tccattcctg 2340
ttattggaga tgaaaataaa aagcaactta tgaaagtaga cattcagatc cagccattac 2400
taacctattc ctttttgggg gaaatctgag cctagctcag aaaaacataa agcaccttga 2460
aaaagacttg gcagcttcct gataaagcgt gctgtgctgt gcagtaggaa cacatcctat 2520
ttattgtgat gttgtggttt tattatctta aactctgttc catacacttg tataaataca 2580
tggatatttt tatgtacaga agtatgtctc ttaaccagtt cacttattgt actctggcaa 2640
tttaaaagaa aatcagtaaa atattttgct tgtaaaatgc ttaatatcgt gcctaggtta 2700
tgtggtgact atttgaatca aaaatgtatt gaatcatcaa ataaaagaat gtggctattt 2760
tggggagaaa aaaaaaaaa 2779

<210> 9
<211> 103

<212> PRT
<213> Homo sapiens

<300>
<308> Genbank ID No.: g1536902

<400> 9
Met Ala Ser Arg Trp Ala Val Gln Leu Leu Leu Val Ala Ala Trp
1 5 10 15
Ser Met Gly Cys Gly Glu Ala Leu Lys Cys Tyr Thr Cys Lys Glu
20 25 30
Pro Met Thr Ser Ala Ser Cys Arg Thr Ile Thr Arg Cys Lys Pro
35 40 45
Glu Asp Thr Ala Cys Met Thr Thr Leu Val Thr Val Glu Ala Glu
50 55 60
Tyr Pro Phe Asn Gln Ser Pro Val Val Thr Arg Ser Cys Ser Ser
65 70 75
Ser Cys Val Ala Thr Asp Pro Asp Ser Ile Gly Ala Ala His Leu
80 85 90
Ile Phe Cys Cys Phe Arg Asp Leu Cys Asn Ser Glu Leu
95 100

<210> 10
<211> 150
<212> PRT
<213> Homo sapiens

<300>
<308> Genbank ID No.: g2388627

<400> 10
Met Asn Leu Trp Leu Leu Ala Cys Leu Val Ala Gly Phe Leu Gly
1 5 10 15
Ala Trp Ala Pro Ala Val His Thr Gln Gly Val Phe Glu Asp Cys
20 25 30
Cys Leu Ala Tyr His Tyr Pro Ile Gly Trp Ala Val Leu Arg Arg
35 40 45
Ala Trp Thr Tyr Arg Ile Gln Glu Val Ser Gly Ser Cys Asn Leu
50 55 60
Pro Ala Ala Ile Phe Tyr Leu Pro Lys Arg His Arg Lys Val Cys
65 70 75
Gly Asn Pro Lys Ser Arg Glu Val Gln Arg Ala Met Lys Leu Leu
80 85 90
Asp Ala Arg Asn Lys Val Phe Ala Lys Leu His His Asn Met Gln
95 100 105
Thr Phe Gln Ala Gly Pro His Ala Val Lys Lys Leu Ser Ser Gly
110 115 120
Asn Ser Lys Leu Ser Ser Ser Lys Phe Ser Asn Pro Ile Ser Ser
125 130 135
Ser Lys Arg Asn Val Ser Leu Leu Ile Ser Ala Asn Ser Gly Leu
140 145 150

<210> 11
<211> 343
<212> PRT
<213> Gallus gallus

<300>
<308> Genbank ID No.: g853834

<400> 11

```
Met Leu Asn Gln Arg Ile His Pro Gly Met Leu Val Leu Leu Met
  1               5                  10                  15
Phe Leu Tyr His Phe Met Glu Asp His Thr Ala Gln Ala Gly Asn
                20                  25                  30
Cys Trp Leu Arg Gln Ala Arg Asn Gly Arg Cys Gln Val Leu Tyr
                35                  40                  45
Lys Thr Asp Leu Ser Lys Glu Glu Cys Cys Lys Ser Gly Arg Leu
                50                  55                  60
Thr Thr Ser Trp Thr Glu Glu Asp Val Asn Asp Asn Thr Leu Phe
                65                  70                  75
Lys Trp Met Ile Phe Asn Gly Gly Ala Pro Asn Cys Ile Pro Cys
                80                  85                  90
Lys Glu Thr Cys Glu Asn Val Asp Cys Gly Pro Gly Lys Lys Cys
                95                 100                 105
Lys Met Asn Lys Lys Asn Lys Pro Arg Cys Val Cys Ala Pro Asp
               110                 115                 120
Cys Ser Asn Ile Thr Trp Lys Gly Pro Val Cys Gly Leu Asp Gly
               125                 130                 135
Lys Thr Tyr Arg Asn Glu Cys Ala Leu Leu Lys Ala Arg Cys Lys
               140                 145                 150
Glu Gln Pro Glu Leu Glu Val Gln Tyr Gln Gly Lys Cys Lys Lys
               155                 160                 165
Thr Cys Arg Asp Val Leu Cys Pro Gly Ser Ser Thr Cys Val Val
               170                 175                 180
Asp Gln Thr Asn Asn Ala Tyr Cys Val Thr Cys Asn Arg Ile Cys
               185                 190                 195
Pro Glu Pro Thr Ser Pro Glu Gln Tyr Leu Cys Gly Asn Asp Gly
               200                 205                 210
Ile Thr Tyr Ala Ser Ala Cys His Leu Arg Lys Ala Thr Cys Leu
               215                 220                 225
Leu Gly Arg Ser Ile Gly Leu Ala Tyr Glu Gly Lys Cys Ile Lys
               230                 235                 240
Ala Lys Ser Cys Glu Asp Ile Gln Cys Ser Ala Gly Lys Lys Cys
               245                 250                 255
Leu Trp Asp Phe Lys Val Gly Arg Gly Arg Cys Ala Leu Cys Asp
               260                 265                 270
Glu Leu Cys Pro Glu Ser Lys Ser Asp Glu Ala Val Cys Ala Ser
               275                 280                 285
Asp Asn Thr Thr Tyr Pro Ser Glu Cys Ala Met Lys Glu Ala Ala
               290                 295                 300
Cys Ser Met Gly Val Leu Leu Glu Val Lys His Ser Gly Ser Cys
               305                 310                 315
Asn Ser Ile Asn Glu Asp Pro Glu Glu Glu Glu Glu Asp Glu Asp
               320                 325                 330
Gln Asp Tyr Ser Phe Pro Ile Ser Ser Ile Leu Glu Trp
               335                 340
```

<210> 12
<211> 730
<212> PRT
<213> Homo sapiens

<300>
<308> Genbank ID No.: g179500

<400> 12
Met Pro Gly Val Ala Arg Leu Pro Leu Leu Leu Gly Leu Leu Leu
1 5 10 15
Leu Pro Arg Pro Gly Arg Pro Leu Asp Leu Ala Asp Tyr Thr Tyr
20 25 30
Asp Leu Ala Glu Glu Asp Asp Ser Glu Pro Leu Asn Tyr Lys Asp
35 40 45
Pro Cys Lys Ala Ala Ala Phe Leu Gly Asp Ile Ala Leu Asp Glu
50 55 60
Glu Asp Leu Arg Ala Phe Gln Val Gln Gln Ala Val Asp Leu Arg
65 70 75
Arg His Thr Ala Arg Lys Ser Ser Ile Lys Ala Ala Val Pro Gly
80 85 90
Asn Thr Ser Thr Pro Ser Cys Gln Ser Thr Asn Gly Gln Pro Gln
95 100 105
Arg Gly Ala Cys Gly Arg Trp Arg Gly Arg Ser Arg Ser Arg Arg
110 115 120
Ala Ala Thr Ser Arg Pro Glu Arg Val Trp Pro Asp Gly Val Ile
125 130 135
Pro Phe Val Ile Gly Gly Asn Phe Thr Gly Ser Gln Arg Ala Val
140 145 150
Phe Arg Gln Ala Met Arg His Trp Glu Lys His Thr Cys Val Thr
155 160 165
Phe Leu Glu Arg Thr Asp Glu Asp Ser Tyr Ile Val Phe Thr Tyr
170 175 180
Arg Pro Cys Gly Cys Cys Ser Tyr Val Gly Arg Arg Gly Gly Gly
185 190 195
Pro Gln Ala Ile Ser Ile Gly Lys Asn Cys Asp Lys Phe Gly Ile
200 205 210
Val Val His Glu Leu Gly His Val Val Gly Phe Trp His Glu His
215 220 225
Thr Arg Pro Asp Arg Asp Arg His Val Ser Ile Val Arg Glu Asn
230 235 240
Ile Gln Pro Gly Gln Glu Tyr Asn Phe Leu Lys Met Glu Pro Gln
245 250 255
Glu Val Glu Ser Leu Gly Glu Thr Tyr Asp Phe Asp Ser Ile Met
260 265 270
His Tyr Ala Arg Asn Thr Phe Ser Arg Gly Ile Phe Leu Asp Thr
275 280 285
Ile Val Pro Lys Tyr Glu Val Asn Gly Val Lys Pro Pro Ile Gly
290 295 300
Gln Arg Thr Arg Leu Ser Lys Gly Asp Ile Ala Gln Ala Arg Lys
305 310 315
Leu Tyr Lys Cys Pro Ala Cys Gly Glu Thr Leu Gln Asp Ser Thr
320 325 330
Gly Asn Phe Ser Ser Pro Glu Tyr Pro Asn Gly Tyr Ser Ala His

```
                335                 340                 345
Met His Cys Val Trp Arg Ile Ser Val Thr Pro Gly Glu Lys Ile
                350                 355                 360
Ile Leu Asn Phe Thr Ser Leu Asp Leu Tyr Arg Ser Arg Leu Cys
                365                 370                 375
Trp Tyr Asp Tyr Val Glu Val Arg Asp Gly Phe Trp Arg Lys Ala
                380                 385                 390
Pro Leu Arg Gly Arg Phe Cys Gly Ser Lys Leu Pro Glu Pro Ile
                395                 400                 405
Val Ser Thr Asp Ser Arg Leu Trp Val Glu Phe Arg Ser Ser Ser
                410                 415                 420
Asn Trp Val Gly Lys Gly Phe Phe Ala Val Tyr Glu Ala Ile Cys
                425                 430                 435
Gly Gly Asp Val Lys Lys Asp Tyr Gly His Ile Gln Ser Pro Asn
                440                 445                 450
Tyr Pro Asp Asp Tyr Arg Pro Ser Lys Val Cys Ile Trp Arg Ile
                455                 460                 465
Gln Val Ser Glu Gly Phe His Val Gly Leu Thr Phe Gln Ser Phe
                470                 475                 480
Glu Ile Glu Arg His Asp Ser Cys Ala Tyr Asp Tyr Leu Glu Val
                485                 490                 495
Arg Asp Gly His Ser Glu Ser Ser Thr Leu Ile Gly Arg Tyr Cys
                500                 505                 510
Gly Tyr Glu Lys Pro Asp Asp Ile Lys Ser Thr Ser Ser Arg Leu
                515                 520                 525
Trp Leu Lys Phe Val Ser Asp Gly Ser Ile Asn Lys Ala Gly Phe
                530                 535                 540
Ala Val Asn Phe Phe Lys Glu Val Asp Glu Cys Ser Arg Pro Asn
                545                 550                 555
Arg Gly Gly Cys Glu Gln Arg Cys Leu Asn Thr Leu Gly Ser Tyr
                560                 565                 570
Lys Cys Ser Cys Asp Pro Gly Tyr Glu Leu Ala Pro Asp Lys Arg
                575                 580                 585
Arg Cys Glu Ala Ala Cys Gly Gly Phe Leu Thr Lys Leu Asn Gly
                590                 595                 600
Ser Ile Thr Ser Pro Gly Trp Pro Lys Glu Tyr Pro Pro Asn Lys
                605                 610                 615
Asn Cys Ile Trp Gln Leu Val Ala Pro Thr Gln Tyr Arg Ile Ser
                620                 625                 630
Leu Gln Phe Asp Phe Phe Glu Thr Glu Gly Asn Asp Val Cys Lys
                635                 640                 645
Tyr Asp Phe Val Glu Val Arg Ser Gly Leu Thr Ala Asp Ser Lys
                650                 655                 660
Leu His Gly Lys Phe Cys Gly Ser Glu Lys Pro Glu Val Ile Thr
                665                 670                 675
Ser Gln Tyr Asn Asn Met Arg Val Glu Phe Lys Ser Asp Asn Thr
                680                 685                 690
Val Ser Lys Lys Gly Phe Lys Ala His Phe Phe Ser Glu Lys Arg
                695                 700                 705
Pro Ala Leu Gln Pro Pro Arg Gly Arg Pro His Gln Leu Lys Phe
                710                 715                 720
Arg Val Gln Lys Arg Asn Arg Thr Pro Gln
                725                 730
```